How Karabo Supports Science Users

Vincenzo Timmel, Simon Felix

SKACH

ETH zürich

University of Applied Sciences and Arts Northwestern Switzerland
School of Engineering

❶ Install Karabo

https://i4ds.github.io/Karabo-Pipeline/installation_user.html

Karabo

Search docs

USERS

- Installation (User)
  - System Requirements
  - Install Karabo
  - Update to the current Karabo version
  - Additional Notes and Troubleshooting
- Containers
- Examples

SIMULATION

- karabo.simulation

IMAGING

- karabo.imaging

/ Installation (User) View page source

Installation (User)

System Requirements

- Linux or Windows with WSL. For macOS we recommend you use Docker.
- 8GB RAM
- 10GB disk space
- GPU-acceleration requires proprietary nVidia drivers/CUDA >= 11.7

Install Karabo

The following steps will install Karabo and its prerequisites (miniconda):

```
wget https://repo.anaconda.com/miniconda/Miniconda3-py39_22.11.1-1-Linux-x86_64.sh
bash Miniconda3-py39_22.11.1-1-Linux-x86_64.sh -b
source ~/miniconda3/bin/activate
conda init bash
conda install -y -n base conda-libmamba-solver
conda config --set solver libmamba
conda update -y -n base -c defaults conda
conda create -y -n karabo-env python=3.9
conda activate karabo-env
```

❷ Try Karabo with 30GB of point sources

```
[8]  sky = sky.filter_by_radius(0, 3, phase_center[0], phase_center[1])

...  Canceled future for execute_request message before replies were done

     The Kernel crashed while executing code in the the current cell or a previous ce
```

❸ Report Problem

https://github.com/i4Ds/Karabo-Pipeline/issues/329

Sign up

i4Ds / Karabo-Pipeline Public

Notifications | Fork 4 | Star 6

Code | Issues 79 | Pull requests 3 | Actions | Projects | Wiki | Security | Insights

Memory issues if the input catalogs for the SkyModel are too large #329

New issue

Closed jejestern opened this issue on Feb 22 · 1 comment

jejestern commented on Feb 22 Collaborator

My input catalogs for the SkyModel are hd5f files which ultimately will be quite large. Maybe it would be better to not convert the catalog into a numpy array if this is an option? Or is there other ways to deal with this?

jejestern added enhancement cscs prio-med labels on Feb 22

deiruch added the user-request label on Feb 23

deiruch assigned kenfus on Feb 23

Assignees: kenfus

Labels: cscs enhancement prio-high user-request

Projects: None yet

Milestone

Post your problem here https://github.com/i4Ds/Karabo-Pipeline/issues/new

❹ We Implement `Xarray DataArrays`

- Read in arbitrarily big skies in Karabo
- `.h5` files are supported
  What else should be added?

```
# Get GLEAM Survey Sky
phase_center = [21.44213503, -30.70729488]
sky = measure_memory(SkyModel.get_BATTYE_sky)
```

✓ 55.4s

```
<class 'xarray.core.dataarray.DataArray'>
Memory usage: 12.08203125 MiB
```

```
sky.sources
```

✓ 0.3s

xarray.DataArray 'array-ce6da31f37610cbc8f82e9939188d172' (source_name: 1119647576, columns: 14)

	Array	Chunk
Bytes	116.79 GiB	128.00 MiB
Shape	(1119647576, 14)	(1198372, 14)
Dask graph	935 chunks in 19 graph layers	
Data type	float64 numpy.ndarray	

Sky by Jennifer Studer

❺ We Improve the Usability

All Karabo functions still work:
Indexing, Slicing, Filtering, …

```
sky_filtered = measure_memory(sky.filter_by_radius_euclidean_flat_approximation, 0, 5,
✓ 2m 29.2s
Calculating distances
Calculating Mask
Filtering sources
Rechunking sky
Memory usage: 8721.859375 MiB
```

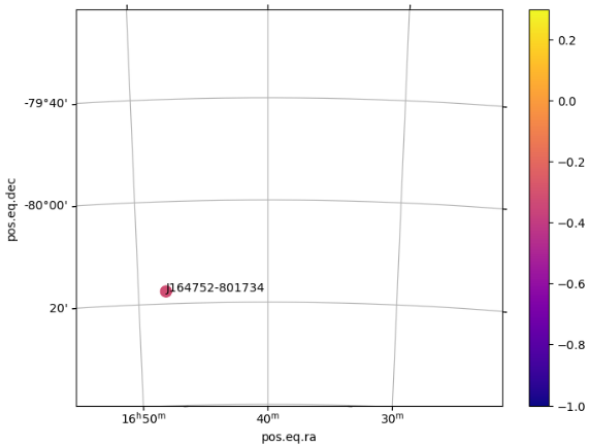
```
sky_filtered.setup_default_wcs(phase_center=phase_center)
```

```
sky_filtered.sources
✓ 0.0s
xarray.DataArray 'array-5ddea2bd6de26655759a105925534a5c' (source_name: 6334002, columns: 14)
```

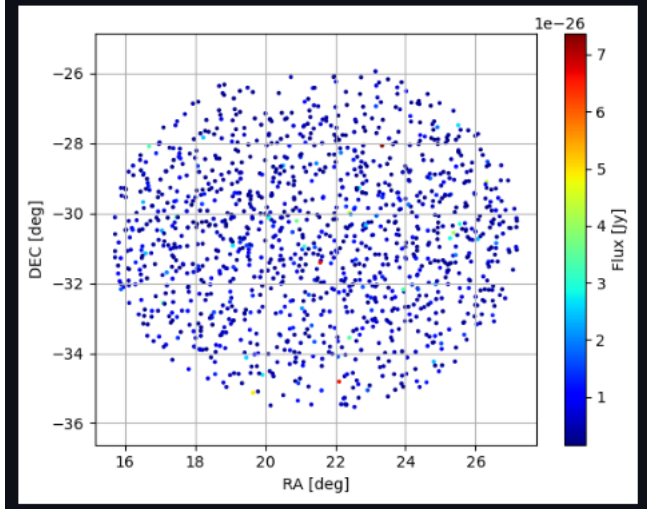
	Array	Chunk
Bytes	676.54 MiB	128.00 MiB
Shape	(6334002, 14)	(1198372, 14)
Dask graph	6 chunks in 21 graph layers	
Data type	float64 numpy.ndarray	

```
sky_specific_name = sky.sources[sky.sources.source_name == "J164752-801734"]
✓ 0.0s
```

```
SkyModel(sky_specific_name).explore_sky(
    phase_center=phase_center,
    figsize=(8, 6),
    xlim=(254, 246),  # RA-lim
    ylim=(-81, -79),  # DEC-lim
    with_labels=True,
    s=80,
    vmin=-1,
    vmax=0.3,
)
✓ 1.3s
```




```
sky_filtered.explore_sky(
    phase_center,
    s=3, cmap='jet', cbar_label='Flux [Jy]', cfun=None, wcs_enabled=False, xlabel='RA
)
✓ 1m 12.8s
```

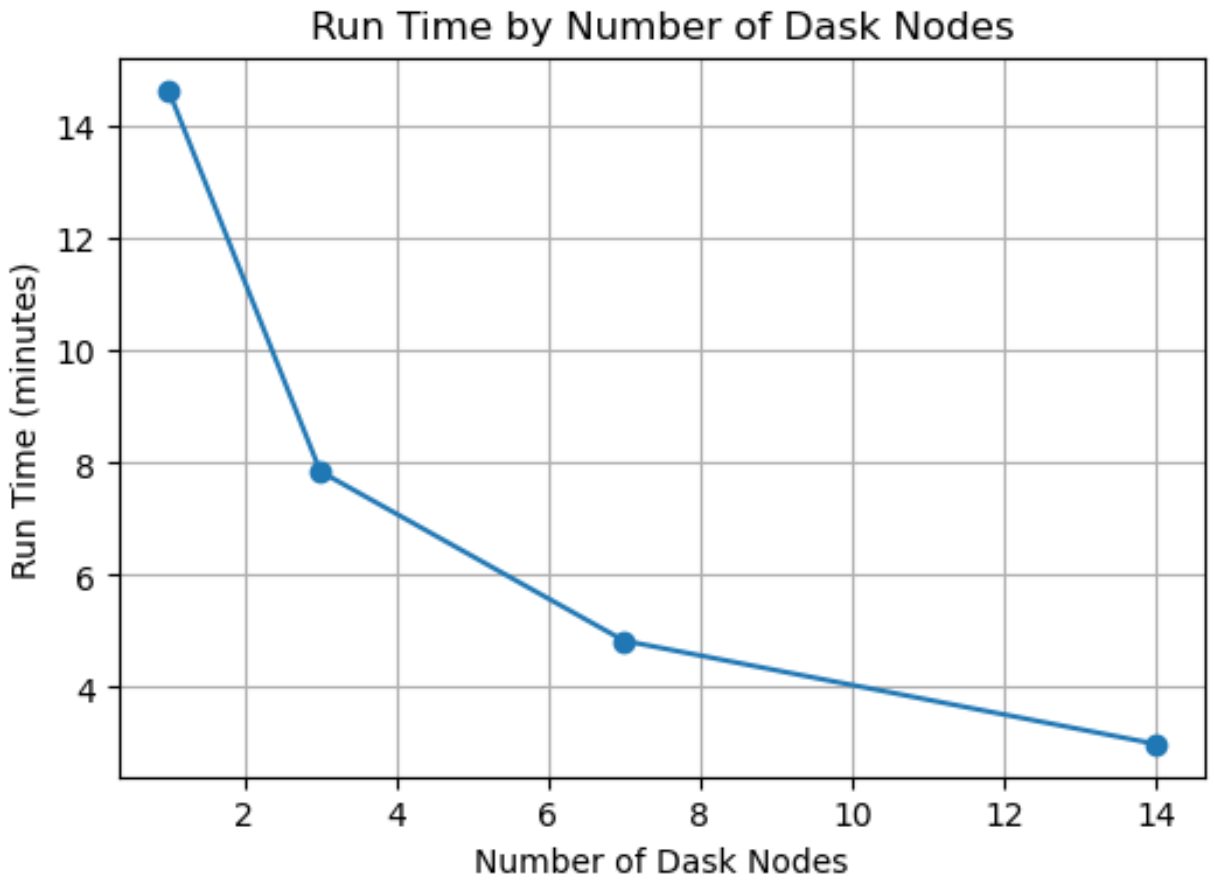


“Automagic” Configuration

- Karabo automatically creates a Dask cluster to distribute work to SLURM nodes
- Automatically uses available cluster resources (e.g. CSCS nodes)
- Currently, OSKAR and RASCIL workloads are distributed to nodes.

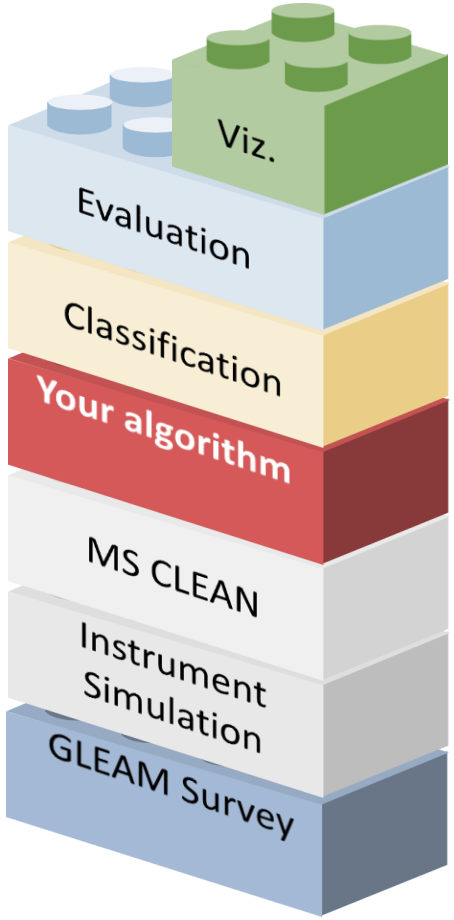
```python

interferometer_sim = InterferometerSimulation(
    channel_bandwidth_hz=1e6,
    use_gpus=True,
    use_dask=True,
    split_observation_by_channels=True,
    n_split_channels="each",
)

visibility = interferometer_sim.run_simulation(
    ...
)
```

“Automagic” Speedup



OSKAR inference simulation of 20'999 point sources
100 channels, parallelized

The Vision



Karabo becomes **the** repository for Swiss SKA software

Useable

Interoperable

Managed